LETTER

doi:10.1038/nature10531

# Somatic retrotransposition alters the genetic landscape of the human brain

J. Kenneth Baillie[1]*, Mark W. Barnett[1]*, Kyle R. Upton[1]*, Daniel J. Gerhardt[2], Todd A. Richmond[2], Fioravante De Sapio[1], Paul Brennan[3], Patrizia Rizzu[4], Sarah Smith[1], Mark Fell[1], Richard T. Talbot[1], Stefano Gustincich[5], Thomas C. Freeman[1], John S. Mattick[6], David A. Hume[1], Peter Heutink[4], Piero Carninci[7], Jeffrey A. Jeddeloh[2] & Geoffrey J. Faulkner[1]

**Retrotransposons are mobile genetic elements that use a germline 'copy-and-paste' mechanism to spread throughout metazoan genomes[1]. At least 50 per cent of the human genome is derived from retrotransposons, with three active families (L1, *Alu* and SVA) associated with insertional mutagenesis and disease[2,3]. Epigenetic and post-transcriptional suppression block retrotransposition in somatic cells[4,5], excluding early embryo development and some malignancies[6,7]. Recent reports of L1 expression[8,9] and copy number variation[10,11] in the human brain suggest that L1 mobilization may also occur during later development. However, the corresponding integration sites have not been mapped. Here we apply a high-throughput method to identify numerous L1, *Alu* and SVA germline mutations, as well as 7,743 putative somatic L1 insertions, in the hippocampus and caudate nucleus of three individuals. Surprisingly, we also found 13,692 somatic *Alu* insertions and 1,350 SVA insertions. Our results demonstrate that retrotransposons mobilize to protein-coding genes differentially expressed and active in the brain. Thus, somatic genome mosaicism driven by retrotransposition may reshape the genetic circuitry that underpins normal and abnormal neurobiological processes.**

Malignancy and ageing are commonly associated with the accumulation of deleterious mutations that lead to loss of function, cell death or uncontrolled growth. Retrotransposition is strongly mutagenic; an estimated 400 million retrotransposon-derived structural variants are present in the global human population[3] and more than 70 diseases involve heritable and *de novo* retrotransposition events[2]. Presumably for this reason, transposition-competent retrotransposons are heavily methylated and transcriptionally inactivated[4,5]. Nevertheless, substantial somatic L1 retrotransposition has been detected in neural cell lineages[10–12]. Given the complex structural and functional organization of the mammalian brain, its adaptive and regenerative capabilities[13] and the unresolved aetiology of many neurobiological disorders, these somatic insertions could be of great importance[14].

One explanation for the observed transpositional activity in the brain may be that the L1 promoter is transiently released from epigenetic suppression during neurogenesis[11,12]. Transposition-competent L1 retrotransposons can then repeatedly mobilize to different loci in individual cells and produce somatic mosaicism. Several lines of evidence support this model, including L1 transcription[8,9] and copy number variation (CNV) in brain tissues from human donors of various ages[10,11], as well as mobilization of engineered L1 retrotransposons *in vitro* and in transgenic rodents[10,12]. Importantly, it is not known where somatic L1 insertions occur in the genome, nor, considering that open chromatin is susceptible to L1 integration[15], whether these events disproportionately affect protein-coding loci expressed in the brain.

Mapping the individual retrotransposition events that collectively form a somatic mosaic is challenging owing to the rarity of each mutant allele in a heterogeneous cell population. We therefore developed a high-throughput protocol that we call retrotransposon capture sequencing (RC-seq). First, fragmented genomic DNA was hybridized to custom sequence capture arrays targeting the 5′ and 3′ termini of full-length L1, *Alu* and SVA retrotransposons (Fig. 1a and Supplementary Tables 1 and 2). Immobile ERVK and ERV1 long terminal repeat (LTR) elements were included as negative controls. Second, the captured DNA was deeply sequenced, yielding ~25 million paired-end 101-mer reads per sample (Fig. 1b). Last, read pairs were mapped using a conservative computational pipeline designed to identify known (Fig. 1c) and novel (Fig. 1d and Supplementary Fig. 1a–d) retrotransposon insertions with uniquely mapped read pairs ('diagnostic reads') spanning their termini.

Previous works have equated L1 CNV with somatic mobilization *in vivo*[10,11]. To test this assumption with RC-seq, we first screened five brain subregions taken from three individuals (donors A, B and C) for L1 CNV. A significant ($P < 0.001$) increase was observed in the number of copies of L1 open reading frame 2 (ORF2) present in DNA extracted from the hippocampus of donor C, and a similar, though smaller, increase was observed for donor A (Fig. 2). We then applied RC-seq to the brain regions that showed the highest (hippocampus) and lowest (caudate nucleus) L1 CNV using samples from all three donors, including a technical replicate of caudate nucleus from donor A. A total of 177.4 million RC-seq paired-end reads were generated from seven libraries (Supplementary Table 3). RC-seq achieved deep sequencing coverage of known active retrotransposons, high reproducibility and limited sequence capture bias (Supplementary Results).

Read pairs diagnostic for novel retrotransposon insertions were clustered on the basis of their insertion site, relative orientation and retrotransposon family. A total of 25,229 clusters were produced. Proximal clusters arranged on opposing strands indicated two termini of one insertion and were paired, resulting in a catalogue of 24,540 novel insertions (Supplementary Table 4). As expected, the great majority of these were either L1 (32.2%) or *Alu* (60.9%) (Fig. 3a). To segregate germline mutations from other events, we combined the three largest available catalogues of L1 and *Alu* polymorphisms[6,16,17] as an annotation database and also performed RC-seq on pooled genomic DNA extracted from blood, producing 6,150 clusters (Supplementary Table 5) that were intersected with the existing brain RC-seq clusters. Any brain clusters that contained RC-seq reads from more than one region or individual, overlapped a blood RC-seq cluster or matched a known polymorphism were designated as germline insertions. Overall, 8.4% of *Alu* insertions in the brain were annotated as germline, versus only 1.9% for L1. Nearly all unannotated

[1]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Edinburgh EH25 9RG, UK. [2]Roche NimbleGen, Inc., 500 South Rosa Road, Madison, Wisconsin 53719, USA. [3]Edinburgh Cancer Research Centre, Western General Hospital, Crewe Road South, Edinburgh EH4 2XR, UK. [4]Section of Medical Genomics, Department of Clinical Genetics, VU University Medical Center, Van der Boechorststraat 7, 1081 BT Amsterdam, The Netherlands. [5]Sector of Neurobiology, International School for Advanced Studies (SISSA), via Bonomea 265, 34136 Trieste, Italy. [6]Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland 4072, Australia. [7]RIKEN Yokohama Institute, Omics Science Center, 1-7-22 Suehiro-chô, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan.

*These authors contributed equally to this work.



©2011 Macmillan Publishers Limited. All rights reserved

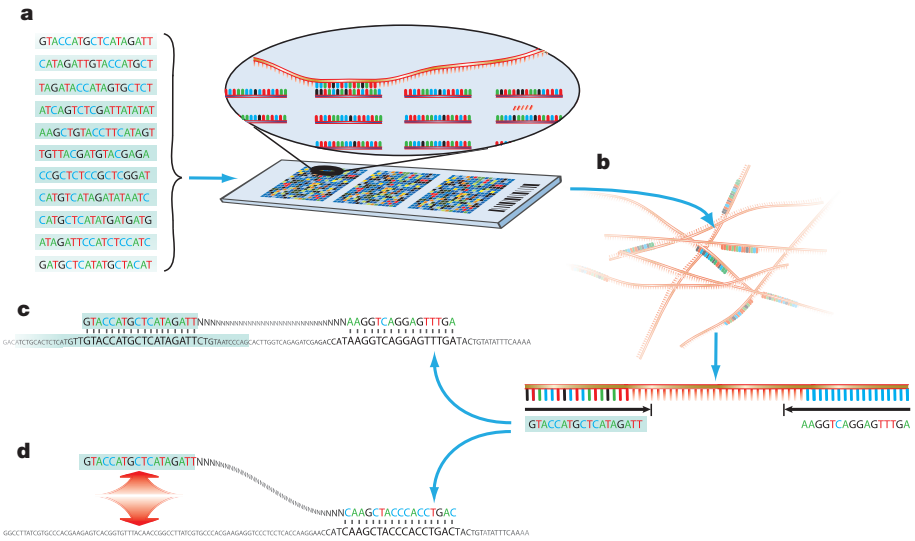


**Figure 1 | Overall RC-seq methodology.** **a**, Retrotransposon capture: sheared genomic DNA is hybridized to custom tiling arrays probing full-length retrotransposons (nucleotides highlighted with light-blue background). **b**, Sequencing: after hybridization, DNA fragments are eluted and analysed with an Illumina sequencer, producing ~$2.5 \times 10^7$ paired-end reads per library that are subsequently aligned to the reference genome. **c**, Reads mapping as a pair to a single locus indicate known retrotransposon insertions. **d**, Unpaired reads where one end maps to a single locus and the other end maps to a distal retrotransposon indicate novel retrotransposition events.

L1 insertions matched fewer than three diagnostic RC-seq reads (Fig. 3b) and were considered potential somatic insertions.

Candidate insertions were validated by PCR amplification and capillary sequencing. Thirty-five germline L1, *Alu*, SVA and LTR insertions were readily confirmed by single-step PCR (Supplementary Table 6). Given low target molecule abundance and the high genomic frequency of the L1 3′ end, we devised a 5′-end nested PCR validation assay for somatic insertions. From 850 and 2,601 full-length (≥90%) L1 and *Alu* insertions, respectively representing 11.0% and 19.0% of the putative somatic insertions found for each family, we selected 29 examples (14 L1 and 15 *Alu*) for validation. Nearly all of the chosen examples were exonic or intronic and were prioritized on the basis of the degree of 5′ truncation, with longer insertions preferred. Optimization of the protocol, combined with substantial input DNA (100 ng), ultimately led to the confirmation of all of the L1 insertions and 12 of the 15 *Alu* insertions (Supplementary Table 7 and Supplementary Fig. 2). Four somatic SVA insertions were also assayed using the same process, and two were confirmed (both from subfamily SVA_F) before the available input material was exhausted.

Repeated attempts to PCR-amplify the corresponding 3′ junctions consistently yielded off-target amplicons, leaving validation based exclusively on 5′ junctions. For this reason, we could not experimentally identify the target-site duplications (TSDs) that are indicative of retrotransposition through target-primed reverse transcription[1]. We propose that the 3′ junctions of insertions validated at their 5′ ends did not amplify efficiently owing to the confounding factors listed above, and to the presence of long polyA tails in on-target amplicons but often not, as we found, in off-target amplicons.

However, TSDs could in some cases be found directly by RC-seq (Supplementary Fig. 1d). An examination of germline insertions sequenced to high depth (≥10 reads) at both their 5′ and their 3′ ends revealed that 43 of 50 (86%) presented TSDs. Owing to their very low abundance—and, therefore, low sequencing coverage—only three putative somatic insertions were detected by at least one RC-seq read at both termini. Two of these insertions (one L1 and one *Alu*) presented TSDs. Despite these and other data strongly supporting retrotransposition as the main cause of somatic mobilization (Supplementary Results), an insufficient number of insertions were sequenced at both ends to determine whether target-primed reverse transcription or an alternative retrotransposition mechanism[18] was primarily responsible.

The somatic origin of each insertion was demonstrated by its presence in one of the assayed brain tissues and its absence from the other, according to RC-seq and PCR results. As illustrative examples, an

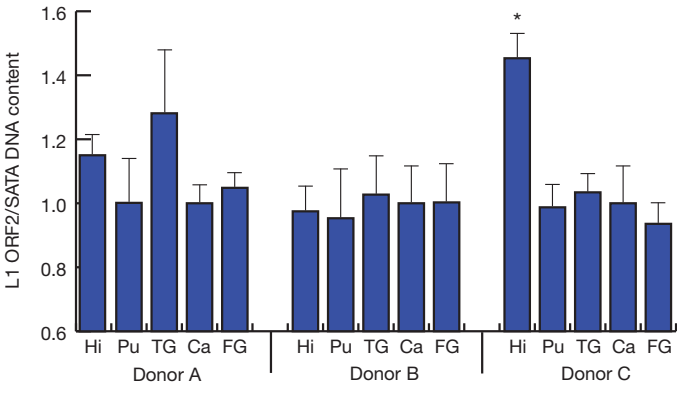


**Figure 2 | Multiplex quantitative PCR confirms L1 CNV in the human brain.** The relative abundance of L1 open reading frame 2 (ORF2) with respect to α-satellite repeats (SATA) was quantified using an existing TaqMan-based approach[10]. Genomic DNA from five brain regions was assayed in three donors (A, B and C). Hi, hippocampus; Pu, putamen; TG, middle temporal gyrus; Ca, caudate nucleus; FG, middle frontal gyrus. Values are normalized to caudate nucleus for each donor. Error bars, s.e.m. *$P < 0.001$ for repeated-measures one-way analysis of variance within each donor, followed by pairwise least-significant-difference *post hoc* tests with Bonferroni correction.

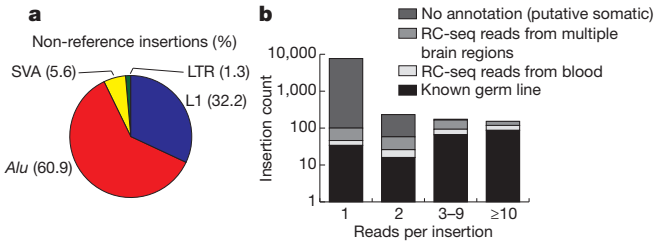


**Figure 3 | Characterization of non-reference genome insertions.** **a**, Proportions of novel insertions found for each family. **b**, Annotation of novel L1 insertions (note logarithmic scale) across all brain libraries. The great majority of insertions detected by fewer than three reads could not be annotated and were considered putative somatic events.



©2011 Macmillan Publishers Limited. All rights reserved

intronic somatic L1 insertion in *HDAC1* is detailed in Fig. 4a, b and an exonic somatic *Alu* insertion in *RAI1* is shown in Fig. 4c, d. These experimental results indicated that insertions detected by RC-seq occurred *in vivo* and did not represent sequencing artefacts.

Donor element annotation revealed that 80.2% of somatic L1 insertions corresponded to the most recently active human L1 subfamilies, L1-Ta and pre-Ta (Supplementary Fig. 3a). The normalized hippocampus/caudate nucleus ratios for somatic L1 insertions were 1.3, 0.5 and 2.2 for donors A, B and C, respectively, paralleling trends from the L1 CNV assay (Fig. 2). Protein-coding loci were disproportionately affected (Supplementary Table 8) relative to random expectation and prior germline frequencies ($P < 0.0001$ for exons and introns, $\chi^2$-test). Pre-existing microarray expression data indicated that genes containing intronic L1 insertions were twice as likely to be differentially overexpressed in the brain as would be randomly expected ($P < 0.0001$, $\chi^2$-test). Key loci were found to contain somatic L1 insertions, including tumour suppressor genes deleted in neuroblastoma and glioma (for example *CAMTA1*), dopamine receptors (*DRD3*) and neurotransmitter transporters (*SLC6A5*, *SLC6A6* and *SLC6A9*). Globally, a gene ontology analysis revealed enrichment for terms relevant to neurogenesis and synaptic function (Supplementary Table 9).

Unlike that for L1, *Alu* retrotransposition has not previously been reported in normal brain cells. However, the L1 transposition machinery is known to *trans*-mobilize *Alu* (ref. 19) and 83.0% of the somatic *Alu* insertions corresponded to the AluY subfamily most active in the human germ line (Supplementary Fig. 3b), making the coincidence of somatic L1 and *Alu* mobilization plausible. On a per-element basis, the observed *Alu* activity was approximately 20-fold lower than that of L1 (Supplementary Results). Thus, it is unlikely that *Alu* CNV would be statistically significant if assayed by TaqMan quantitative PCR[10]. The genomic patterns of *Alu* and L1 insertions were also different; somatic *Alu* insertions were not overrepresented in introns but were even more common than L1 in exons (Supplementary Table 8). *Alu* exonization is a noted cause of genetic disease[2]. Overall, L1, *Alu* and, to a more limited extent, SVA mobilization produced a large number of insertions that affected protein-coding genes.

Our results indicate that somatic L1 and *Alu* mobilization fundamentally alters the genetic landscape of the human brain, and that retrotransposition is the primary mechanism underlying this phenomenon. By contrast with germline activity[6,16], somatic insertions disproportionately impacted protein-coding loci. Germline insertions are rarely found in regions where they generate a deleterious phenotype because such mutations are strongly selected against during evolution. Somatic events, on the other hand, are present for one generation and may affect protein-coding loci in a specific environmental context, perhaps being drawn to open chromatin in transcribed regions[15]. Apart from the obvious effects of exonic insertions, intronic events could act as subtle transcriptional 'rheostats'[20] or as *cis*-regulatory elements[21] akin to the IAP insertion responsible for the viable yellow allele of the agouti gene in the mouse[22].

Several recent studies have catalogued retrotransposon insertions in the human germ line and tumours[6,16,23,24]. Through RC-seq, we have extended these data to the brain and linked somatic retrotransposition to neurobiological genes. For instance, HDAC1 is a genome-wide transcriptional regulator that controls the canonical L1 promoter[4,25] and is implicated in psychiatric disease and tumorigenesis[26]. Another example highlighted here, RAI1, is a transcription factor highly expressed in the brain and previously linked with schizophrenia and Smith–Magenis syndrome[27]. An exonic *Alu* insertion in *RAI1* (Fig. 4c), could therefore have phenotypic consequences.

The hippocampus seems to be predisposed to somatic L1 retrotransposition[10], which is intriguing given that its subgranular zone is a

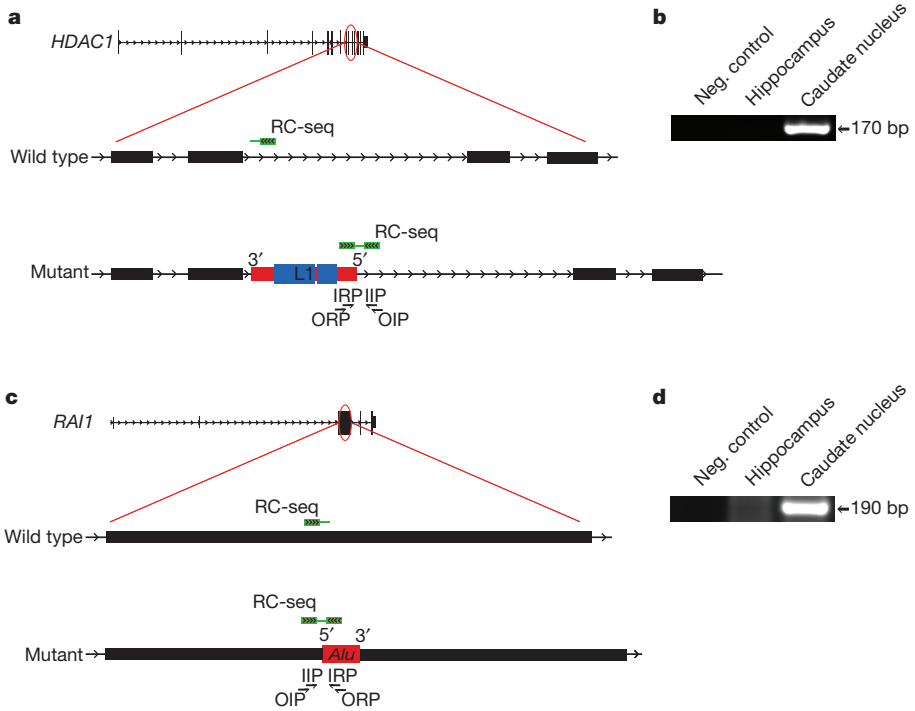


**Figure 4 | Discovery of somatic insertions in *HDAC1* and *RAI1*.** **a**, Alignment of an RC-seq read (green) from donor C caudate nucleus indicated an antisense L1 insertion in intron 9 of *HDAC1*. Nested PCR primers were designed to span the L1 5′ terminus, with an initial reaction combining outside retrotransposon primers (ORPs) and insertion site primers (OIPs) and a second reaction using inside retrotransposon primers (IRP) and insertion site primers (IIP). **b**, Amplification of the nested PCR target, confirmed for specificity by capillary sequencing, was achieved in caudate nucleus but not in hippocampus. Sequencing indicated that the L1 insertion mobilized from chromosome 9 and was accompanied by 5′ transduction. bp, base pairs. **c**, Alignment of an RC-seq read pair from donor A caudate nucleus indicated a sense *Alu* insertion in exon 3, and the coding sequence, of *RAI1*. **d**, As for **b**, amplification of the nested PCR target was achieved in caudate nucleus but not in hippocampus. Sequencing indicated that the *Alu* insertion mobilized from chromosome 4. We note that L1 and *Alu* elements in **a** and **c** are not drawn to scale, that the L1 open reading frames in **a** are coloured in blue and that the untranslated regions of L1 and *Alu* are coloured in red.



©2011 Macmillan Publishers Limited. All rights reserved

main source of adult neurogenesis[13]. This is also consistent with the hypothesis that L1 retrotransposition is related to neural plasticity[14]. Even more intriguing is the possibility that the APOBEC proteins, which are RNA/DNA-editing enzymes that have expanded under strong positive selection in the primate lineage and been shown to control L1 mobility, may modulate somatic retrotransposition in the brain[28].

Mutagenesis due to somatic retrotransposition has obvious tumorigenic potential[29] and may have a role in other diseases and biological processes. For example, deletion of the chromatin-remodelling HDAC1 cofactor MECP2[12,25] leads to increased L1 copy number and may inhibit neuronal maturation in Rett syndrome[30]. Somatic mosaicism could also be a factor in neurological dimorphisms seen among discordant monozygotic twins[14]. Future studies may determine whether the overall frequency of somatic retrotransposition varies considerably between individuals, as suggested by our data and previous experiments[10], and between populations. Ultimately, direct identification of transcripts disrupted by somatic retrotransposition, together with its epigenetic regulation, may provide insights into the molecular processes underlying human cognition, neurodevelopmental disorders and neoplastic transformation.

## METHODS SUMMARY

**Human DNA samples.** Tissues were provided by the Netherlands Brain Bank, Amsterdam. They came from three post-mortem donors with no evidence of neurodegeneration. Pooled human genomic DNA was purchased from Promega.

**TaqMan quantitative PCR.** Quantitative PCR experiments were performed with minor modifications to an earlier approach[10]. Quantification included five technical replicates. For each assay, the ratio of L1 ORF2 to α-satellite repeats was normalized to the ratio obtained for caudate nucleus. Ratios were compared across brain regions by repeated-measures one-way analysis of variance with Bonferroni correction.

**Retrotransposon capture array design.** A NimbleGen Sequence Capture 2.1M Array was customized to contain oligonucleotide probes tiled across the termini of full-length L1, *Alu* and SVA retrotransposons, as well as LTRs intended to act as negative controls. Probes were not filtered for repetitiveness. Typically, eight probes were generated per L1 and SVA retrotransposon and per LTR, and four probes per *Alu* retrotransposon, with a total of 4,885 probes across 875 targeted elements.

**Capture library preparation and sequencing.** DNA sequencing libraries were constructed using an Illumina paired-end kit with substantial modifications (Supplementary Methods). Genomic DNA (2.5 μg) was used for each RC-seq library. Amplification based on ligation-mediated PCR was performed before and after hybridization. The average insert size was ~250 nucelotides. Enrichment was confirmed by quantitative PCR against *Alu*. Sequencing was performed by ARK-Genomics, The Roslin Institute, using an Illumina GAIIx instrument.

**Computational analyses.** Paired-end RC-seq reads were mapped to human genome assembly hg19 using SOAP2. Reads where both ends could be aligned to the genome, but not at the same locus, indicated novel retrotransposon insertions. These alignments were corroborated by BLAT, stringently filtered and clustered. Clusters were annotated using published retrotransposon databases[6,16,17] and the NCBI RefSeq database.

**Received 14 March; accepted 5 September 2011.**
**Published online 30 October 2011.**

1. Kazazian, H. H. Jr. Mobile elements: drivers of genome evolution. *Science* **303,** 1626–1632 (2004).
2. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10,** 691–703 (2009).
3. Xing, J. *et al.* Mobile elements create structural variation: analysis of a complete human genome. *Genome Res.* **19,** 1516–1526 (2009).
4. Garcia-Perez, J. L. *et al.* Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* **466,** 769–773 (2010).
5. Yang, N. & Kazazian, H. H. Jr. L1 retrotransposition is suppressed by endogenously encoded small interfering RNAs in human cultured cells. *Nature Struct. Mol. Biol.* **13,** 763–771 (2006).
6. Iskow, R. C. *et al.* Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141,** 1253–1261 (2010).
7. Kano, H. *et al.* L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* **23,** 1303–1312 (2009).
8. Belancio, V. P., Roy-Engel, A. M., Pochampally, R. R. & Deininger, P. Somatic expression of LINE-1 elements in human tissues. *Nucl. Acids Res.* **38,** 3909–3922 (2010).
9. Faulkner, G. J. *et al.* The regulated retrotransposon transcriptome of mammalian cells. *Nature Genet.* **41,** 563–571 (2009).
10. Coufal, N. G. *et al.* L1 retrotransposition in human neural progenitor cells. *Nature* **460,** 1127–1131 (2009).
11. Muotri, A. R. *et al.* L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468,** 443–446 (2010).
12. Muotri, A. R. *et al.* Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* **435,** 903–910 (2005).
13. Eriksson, P. S. *et al.* Neurogenesis in the adult human hippocampus. *Nature Med.* **4,** 1313–1317 (1998).
14. Singer, T., McConnell, M. J., Marchetto, M. C., Coufal, N. G. & Gage, F. H. LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* **33,** 345–354 (2010).
15. Cost, G. J., Golding, A., Schlissel, M. S. & Boeke, J. D. Target DNA chromatinization modulates nicking by L1 endonuclease. *Nucleic Acids Res.* **29,** 573–577 (2001).
16. Ewing, A. D. & Kazazian, H. H. Jr. High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* **20,** 1262–1270 (2010).
17. Wang, J. *et al.* dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.* **27,** 323–329 (2006).
18. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31,** 159–165 (2002).
19. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nature Genet.* **35,** 41–48 (2003).
20. Han, J. S., Szak, S. T. & Boeke, J. D. Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* **429,** 268–274 (2004).
21. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9,** 397–405 (2008).
22. Morgan, H. D., Sutherland, H. G., Martin, D. I. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nature Genet.* **23,** 314–318 (1999).
23. Beck, C. R. *et al.* LINE-1 retrotransposition activity in human genomes. *Cell* **141,** 1159–1170 (2010).
24. Huang, C. R. *et al.* Mobile interspersed repeats are major structural variants in the human genome. *Cell* **141,** 1171–1182 (2010).
25. Nan, X. *et al.* Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386–389 (1998).
26. Kazantsev, A. G. & Thompson, L. M. Therapeutic application of histone deacetylase inhibitors for central nervous system disorders. *Nature Rev. Drug Discov.* **7,** 854–868 (2008).
27. Slager, R. E., Newton, T. L., Vlangos, C. N., Finucane, B. & Elsea, S. H. Mutations in *RAI1* associated with Smith–Magenis syndrome. *Nature Genet.* **33,** 466–468 (2003).
28. Mattick, J. S. RNA as the substrate for epigenome-environment interactions: RNA guidance of epigenetic processes and the expansion of RNA editing in animals underpins development, phenotypic plasticity, learning, and cognition. *Bioessays* **32,** 548–552 (2010).
29. Miki, Y. *et al.* Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* **52,** 643–645 (1992).
30. Chahrour, M. & Zoghbi, H. Y. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56,** 422–437 (2007).

**Supplementary Information** is linked to the online version of the paper at www.nature.com/nature.

**Acknowledgements** J.K.B. is supported by a Wellcome Trust Clinical Fellowship (090385/Z/09/Z) through the Edinburgh Clinical Academic Track. G.J.F. is funded by an Institute Strategic Programme Grant and a New Investigator Award from the British BBSRC (BB/H005935/1) and a C. J. Martin Overseas Based Biomedical Fellowship from the Australian NHMRC (575585). Human brain tissues were provided by the Netherlands Brain Bank to P.H. with ethical consent for them to be used as described in the study.

**Author Contributions** J.K.B., M.W.B., K.R.U., D.J.G., P.R., S.S., P.C. and G.J.F. designed and performed the experiments. J.K.B., T.A.R., F.D.S. and M.F. conducted the computational analyses. P.B., R.T.T., T.C.F., D.A.H., P.H., P.C., J.A.J. and G.J.F. provided resources. S.G. and J.S.M. contributed to the discussion. J.A.J. and G.J.F. invented RC-seq. G.J.F. directed the study, led the bioinformatic analysis and wrote the manuscript. All authors commented on or contributed to the final manuscript.

**Author Information** RC-seq FASTA sequences for brain and blood have been deposited in the NCBI Sequence Read Archive under accession number SRA024401. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.J.F. (faulknergj@gmail.com).



©2011 Macmillan Publishers Limited. All rights reserved